M.R.P. 931
S.C. 111/190
14th September, 1955

Some statistical aspects of a rain-making experiment

by

A. F. Jenkinson
(Meteorological Office)

METEOROLOGICAL
OFFICE
15 FEB 1956
LERWICK OBSERVATORY
SHETLAND

AIR MINISTRY
METEOROLOGICAL RESEARCH COMMITTEE

At the time of issue this report has not been published. Any author desiring to quote this paper should refer to it as "A paper of the Meteorological Research Committee (London).
M.R.P. No. 931 by Mr. A.F. Jenkinson
A copy is available in the Library of the Meteorological Office."

M.R.P. 931
S.C.III/190
14th September, 1955

AIR MINISTRY
METEOROLOGICAL RESEARCH COMMITTEE

Some statistical aspects of a rain-making experiment

by
A. F. Jenkinson (Meteorological Office)

1. Nature of the experiment

The experiment is one which the Meteorological Office proposes to carry out during the next two or three years and is designed to enable us to assess the effects of cloud seeding from ground based silver iodide generators, located in the Salisbury Plain area, in a particular type of weather situation, viz. cloudy (frontal) situations with south-westerly winds.

The variability of rainfall amount in these situations, although of course less than in other situations in which convection may play a considerable part, is nevertheless considerable, and the ratio of the standard deviation S_x to mean rainfall $\bar{x}$ is of the order of 1.5 if occasions with very slight rainfall are included. (This has also been confirmed from data sent to me by J. S. Sawyer. The data were provided by G. A. Corby). Since these light falls are not of importance to the rainfall total, and since also it is difficult to give any meaning to a proportionate increase of rainfall when the rainfall is small, it is better to consider only occasions which give substantial falls of rain.

To make the selection of occasions objective, and to ensure that the ratio $S_x/\bar{x}$ is as low as possible while retaining an adequate number of occasions, the criterion chosen for selection of an occasion is that there should be a broad south-westerly air stream south of the Humber with layered (frontal) cloud which gives 5mm. or more of rain in the Salisbury area (mean rainfall for four stations). In these situations the ratio $S_x/\bar{x}$ is very nearly 0.66 over a large area of England.

2. Locating the silver iodide plume

At the moment little is known about the diffusion of the plume, and in the cloudy conditions we expect during the seeding operations it may not be possible to make effective use of tracer substances. In the light of present knowledge the best method of locating the plume at the "effective" seeding levels, with temperatures between -3°C. and -15°C., is to adopt the suggestion of R. F. Jones (1954), that the plume at those levels has been carried there by the mean vector wind from the surface to those levels. (In a verbal communication, F. Pasquill has said that this approximation is probably fairly reasonable.)

This gives us the direction in which to look for the plume at the effective seeding levels. Formulae given by O.G. Sutton (1947; 1953, p.287) for diffusion from a point source, indicate that the concentration of silver iodide is a maximum at a height H at a distance D where

$$H = C\, D^{\frac{7}{8}}$$

and C is approximately constant (probably varying from 0.5 to 2.0 times the mean of its value under all conditions). For H in thousands of feet and D in miles, the mean value of C is about 0.2 (R.F. Jones, 1954; O.G. Sutton, 1953, p.292). Corresponding values of H and D for a mean C are given in Table 1.

/TABLE 1.

TABLE 1. CORRESPONDING VALUES OF H AND D

$H(10^3$ ft.)	D(mi.)
2.5	20
5.5	50
10.1	100
14.4	150
18.5	200
22.6	250

Thus, without a knowledge of C_S, the value of C for a particular seeding occasion, there is likely to be some doubt about the horizontal distance of the area effectively seeded. R. F. Jones (1954) suggests that location of a succession of balloons, fitted with radar target and having a slow rate of ascent (100 to 200 ft./min.), released at $\frac{1}{2}$ hour to 1 hour intervals might provide data for an assessment of C_S. These balloons would also give regular wind measurements over approximately the same track as the smoke.

Some knowledge of the diffusion of small particles of silver iodide has been gained from experimental work on the diffusion of small particles of "tracer" substances. The tracer substance most commonly used up to the present time has been zinc sulphide. This has the disadvantage that it can only be used in clear air. Sampling has been carried out at levels up to 4,000 feet. (R.R. Braham, B.K. Seely and W.D. Crozier, 1952).

O. Vittori (1954) has suggested that silver nitrate particles can be emitted along with the silver iodide, and detected by a modification of the Liesegang reaction.

Future work with tracer substances will increase our knowledge of the vertical diffusion of small particles of silver iodide, and it may be that values of C_S will be determined on individual occasions.

3. Vertical and horizontal distribution of the effectively seeded clouds

It may be noted that the assumption in this paper that an effect of seeding is to be detected by statistical methods is equivalent to an assumption that silver iodide smoke diffuses vertically to heights of say 20,000 feet, through cloud and probably through rain. With this assumption let us examine the vertical and horizontal distribution of the effectively seeded clouds.

The vertical diffusion of the silver iodide smoke is depicted in Fig. 1. Since in general wind increases with height, the motion of the silver iodide smoke relative to the cloud formations, is like that of the paths AA^1, BB^1, CC^1. The cloud is above the frontal slope A^1C. The two levels -3°C. and -15°C., between which silver iodide is said to be most effective as a seeding agent, are shown. Then the emissions of silver iodide which first reach the medium cloud layers, AA^1, may well be above the -15°C. level. We may first seed in the "effective" layer at X along BB^1 and the last seeding, through cloud and rain from near the surface, may be supposed to be along CC^1.

Now let us suppose for example that the time between the beginning of movement of silver iodide along BB^1 and CC^1 is seven hours. Then during that period the particle paths will have changed due to the variations of vector wind with time. Fig. 2 shows the horizontal projection of the particle paths of the silver iodide (not streamlines). Then the line OA_1 shows the particle path for the ascent AA^1, OB_1 for BB^1, OC_1 for CC^1. The line $A_1X_1Y_1O$ shows where cloud is first met. X_1X_2 and Y_1Y_2 show where the -15°C. and -3°C. levels are reached. Then the area $X_1Y_1Y_2X_2$, together with any additional area due to the plume having definite width, is the "seeded" area.

On occasions the latitudinal movement of the plume due to the vector change of wind with time may be as much as 20 to 30 miles at a distance from the generators of 100 miles; but if we for the moment consider the plume to be very narrow then the areas seeded during each hour will be those marked in Fig. 2 with the words "First hour", "Second hour" etc., corresponding only to the vertical cloud structures marked "First hour" etc. in Fig. 1. The plume width from a single generator, from the analysis of R. F. Jones (1954), is effectively of the order of 5 to 10 miles, and this width is only slightly increased even by large (ten-fold) increases of the rate of production of silver iodide particles by the generator.

If there is no change of vector wind with time, and we have a single generator, then all the vertical cloud structures marked first hour, second hour etc. will pass over the same horizontal plume width of 5 to 10 miles. It will be seen, if we have a single generator, that as the latitudinal distance which the plume moves due to vector variation of wind with time increases, so does the relative efficiency of seeding of any horizontal area decrease, defining efficiency of seeding as the proportion of the cloud structures which pass over the area; and with a latitudinal time-movement of the plume of 30 miles (at a distance of 100 miles) and a single generator with plume width of 10 miles, each part of the horizontal extent of the plume receives one third or less of the cloud structure, while if the multiple generator plume width is 30 miles there will be an area 30 miles wide over which four sevenths or more of the cloud structure passed, and an area 20 miles wide over which five sevenths or more of the cloud structure passed. Fig. 3 shows what parts of the cloud structure would pass over the various horizontal widths for a latitudinal time-movement of the plume of 30 miles, and plume widths of 10, 20 and 30 miles.

It is evident that to ensure a high seeding efficiency over an area 20 miles wide it is necessary to have a line of generators, say five, over a width of 20 to 25 miles, perpendicular to the S.W. direction. This will also greatly increase the probability that, even with some uncertainty in the angular direction of the plume, we would be able to locate a substantial area which has been efficiently seeded.

4. Statistical analysis

If $y = x/\bar{x}$ where x is a random variable with expected value ξ and standard deviation σ such that $\sigma/\xi = \alpha$ and $\bar{x}$ and S_x are the mean value and standard deviation of x determined from N values, then to order $1/N$ the expected value and variance of y are

$$\left.\begin{aligned} \mathcal{E}(y) &\simeq 1+(\alpha^2/N) \\ \sigma^2(y) &\simeq \alpha^2\left[1+\frac{1+3\alpha^2}{N}\right] \end{aligned}\right\} \quad \text{- - - - - -} \qquad (1)$$

(N. L. Johnson and H. Tetley, 1949, Vol. I, 7.38 and 7.39)

If x is the rainfall of a small area for a particular experimental occasion, then if $\bar{x}$ is the mean derived from over 200 occasions, and we note that $\alpha \simeq 0.66$, then

$$\left.\begin{aligned} \mathcal{E}(y) &\simeq 1.002 \\ \sigma^2(y) &\simeq 1.01\alpha^2 \end{aligned}\right\} \quad \text{- - - - - -} \qquad (2)$$

It is proposed to estimate the mean and standard deviation of the falls of rain in small areas for the selected occasions during the period 1945-54. There are about 20 to 25 occasions a year, and we should have 200 to 250 occasions for our estimate of mean, standard deviation and the ratio α.

/But

But for the purpose of the present preliminary analysis, the rainfall for each of the 92 occasions during the 4-year period 1951-54 has been extracted from the twice-daily readings given in the Daily Weather Report. Rainfalls for West Malling were extracted from M.S. data in the Meteorological Office, M.O.3. The area was divided into 8 sub-areas, A, B, C, D, E, F, G, H. These are shown in Fig. 4. The stations used to give the mean rainfall on each occasion are marked. For each sub-area the mean rainfall for the 92 occasions was computed, and also the ratio $S_x/\bar{x}$, approximately α. The value of $S_x/\bar{x}$ did not vary much from area to area (see Table 5) and we are justified in using a mean value of 0.66 for the whole area. We may be able to assess a more accurate value for α when rainfall for all occasions during 1945-54 is analysed.

Because of the correlation between the rainfall of one sub-area and the rainfall in a neighbouring area, we can estimate the proportion $y = (x/\bar{x})$ received in any area from the proportion $y_1 = (x_1/\bar{x}_1)$ received in a neighbouring area. The estimate y^1 is given by

$$y^1 - \bar{y} = b(y_1 - \bar{y}_1) \qquad (3)$$

where $\bar{y}$, $\bar{y}_1$ are the mean values of y and y_1 in the sample of N, and

$$b = r \frac{S_y}{S_{y_1}} \qquad (4)$$

where r is the correlation coefficient between y and y_1, and S_y, S_{y_1} are the standard deviations of y and y_1, found from the sample. The residual variance S_Y^2 of the differences $Y = y - y^1$ is given by

$$S_Y^2 = S_y^2 (1 - r^2) \qquad (5)$$

It may be noted that r is also the correlation coefficient between x and x_1, and

$$S_y = (S_x/\bar{x}) \qquad S_{y_1} = (S_{x_1}/\bar{x}_1)$$

By using a combination of areas, we can estimate y by a linear regression in terms of the several ratios $y_i (i = 1,2,3....)$ and the residual variance S_Y^2 will be given by

$$S_Y^2 = S_y^2 (1 - R^2) \qquad (6)$$

where R is the multiple correlation coefficient between y and the y_i.

If we contemplate using a ten-year period, with N situations, to estimate the regression, and a three-year period with n situations to test the seeding effects, N and n will be of the order of 230 and 69 respectively.

Then if $\overline{Y_n}$ is the mean value for the n occasions of the differences y and y^1 based on a simple regression, then we can test the null-hypothesis that seeding has not increased rainfall by means of the t-variable

$$\frac{\overline{Y_n}}{\left[S_y^2 (1 - r^2)\left(\frac{1}{n} + \frac{1}{N} + C_{11}\overline{y_{1n}}^2\right)\right]^{1/2}} \qquad (7)$$

for approximately 300 degrees of freedom.

/$\overline{y_{1n}}$

$\overline{y_{1n}}$ is the mean of the n values of $(y_1 - \overline{y_1})$ and C_{11} is the Gaussian coefficient. The term $C_{11}\overline{y_{1n}}^2$ is unimportant with these sample sizes.

Strictly, the t-test can only be applied if the population of Y's, whose expected value is 0, is normally distributed, but the test can be satisfactorily applied if the Y's are reasonably near normal.

The t-test may be extended to a multiple regression, with multiple correlation coefficient R, replacing r by R in expression (7) and $C_{11}\overline{y_{1n}}^2$ by $\sum_{ij} C_{ij}\overline{y_{in}}\,\overline{y_{jn}}$ where i, j take values 1, 2, ... N. This sum is also unimportant with these sample sizes. Multiple regressions are however not easy to compute, and for the purpose of this assessment we can use a simpler approach.

When using a simple correlation, y estimated from y_1, then the best estimation line for y, taking S_y and S_{y_1} to have the same value, is

$$y = ry_1 + (1 - r) \qquad (8).$$

But if we were to consider the best estimation line for y_1 from y this is

$$y_1 = ry + (1 - r) \qquad (9)$$

So, if a straight line really represents a good fit to the data of y and y_1, that is if r is high, there seems, from the physical aspect, much to be said for using the common estimator $y = y_1$. The variance of $(y - y_1) = 2(1-r)\,S_y^2$ and we see from Table 2 that for $r \geq 0.8$ there is little difference between $\sqrt{1 - r^2}$ and $\sqrt{2(1 - r)}$.

TABLE 2

CORRESPONDING VALUES OF $\sqrt{1 - r^2}$ and $\sqrt{2(1 - r)}$

r	$\sqrt{1 - r^2}$	$\sqrt{2(1 - r)}$
0.80	0.600	0.633
0.85	0.527	0.548
0.90	0.436	0.447

If we extend this method and estimate y by

$$y^1 = \frac{1}{m}(y_1 + y_2 + \ldots\ y_m)$$

then $$Y = y - \frac{1}{m}(y_1 + y_2 + \ldots\ y_m)$$

If $S_y^2 = S_{y_1}^2 = S_y^2 = \ldots\ a^2$ (very nearly equal to a^2)

$$S_Y^2 = a^2\left\{(1 + \frac{1}{m})(1 - \overline{r_{ij}}) - 2(\overline{r_i} - \overline{r_{ij}})\right\} \qquad (10)$$

where r_i is the correlation between y and y_i

r_{ij} is the correlation between y_i and y_j

Figs. 5 to 12 show the correlations between each of the sub-areas and the remainder of the area. It will be apparent from the variation of these correlations that around each seeded area it should be possible to find four neighbouring areas each with a correlation of 0.70 or more with the seeded area, but with a smaller average degree of correlation among themselves if they are chosen two each on opposite sides of the seeded area. Table 3 gives the values of $\sqrt{(1 + \frac{1}{m})(1 - \overline{r_{ij}}) - 2(\overline{r_i} - \overline{r_{ij}})}$

/for

for corresponding values of $\overline{r_i}$ and $\overline{r_{ij}}$.

TABLE 3. VALUES OF $\sqrt{(1 + \frac{1}{m})(1 - \overline{r_{ij}}) - 2(\overline{r_i} - \overline{r_{ij}})}$

$\overline{r_i}$ \ $\overline{r_{ij}}$	0.60	0.65	0.70	0.75
0.70	0.55	0.58		
0.75		0.49	0.52	
0.80			0.41	0.46

A value for $\sqrt{(1 + \frac{1}{m})(1 - \overline{r_{ij}}) - 2(\overline{r_i} - \overline{r_{ij}})}$ of 0.60 or less indicates a multiple regression coefficient of 0.80 or more.

We may thus reasonably expect to be able to find for each seeded area a group of areas which have a multiple correlation coefficient with the seeded area of 0.80 to 0.85, giving a value for S_Y of 0.60 a or less.

Even with the rather widely spaced areas used in this analysis, this estimation is confirmed. In estimating the rainfall of area E from rainfalls of G, F, W and H, $\overline{r_i} = 0.69$, $\overline{r_{ij}} = 0.60$ and $S_Y = 0.56$ a.
In estimating the rainfall of area C from rainfalls of A, B, X and W, $\overline{r_i} = 0.71$, $\overline{r_{ij}} = 0.56$, $S_Y = 0.50$ a.

If we take a = 0.66 and we use the (maximum) value $S_Y = 0.60$ a $\simeq 0.40$ for the purpose of this analysis we can ignore the varying locations of the seeded area and its concomitant regression areas, and we can use the t-test (effectively the normal since $n + N \simeq 300$) to test the null hypothesis that there is no increase in rainfall due to seeding. The value of t will be effectively

$$\frac{\overline{Y_n}}{0.40\left(\frac{1}{69} + \frac{1}{230}\right)^{1/2}} = \frac{\overline{Y_n}}{0.055} \qquad (11)$$

Thus, given the answer $\overline{Y_n}$ for the experiment with n seedings, we can give the conclusion:

If $\overline{Y_n} > (2\frac{1}{3} \times 0.055) = 0.13$ then there is a probability of 0.99 that seeding has produced a real increase in rainfall.

So if the experiment reveals that the rainfall in the seeded area is greater by 13 per cent. than would be predicted from the rainfalls of neighbouring non-seeded areas we can say that there is only 1 chance in 100 of this increase having occurred by chance variation of rainfall. If the experiment reveals a 16 per cent. greater value than predicted from neighbouring non-seeded areas we can say that there is only 1 chance in 1000 of this increase having occurred by chance variation of rainfall. If the experiment were to go on for five years instead of three, the respective figures for 0.01 and 0.001 probability levels would be 10 per cent. and 13 per cent.

5. Illustrative example

To make the statistical procedure more clear, a practical analysis was carried out.

For the period 1951-54, the rainfall of every synoptic situation of the

required type was analysed, using the 12-hourly rainfall amounts given in the British D.W.R's. The choice of synoptic situations was objective; the requirement was that there should be a broad south-west stream, south of the Humber, (upper winds from 3,000 feet to the 0°F. level from direction 190°-270°, with some backing of the surface wind permissible) with layered frontal cloud. It was further required that there should have been 5mm. or more of rain at Boscombe Down. (In the actual experiment, of course, this kind of selective choice would be made in retrospect by the analyst. Emission of silver iodide smoke would no doubt be made on doubtful occasions).

The rainfall stations were divided into area groups. These areas, and the rainfall stations, are shown in Fig. 4.

There were 92 selected occasions during the period 1951-54. The seasonal distribution is shown in Table 4.

TABLE 4. DISTRIBUTION OF "SEEDING" OCCASIONS DURING 1951-54

	1951	1952	1953	1954	Total	Mean
November	5	0	2	5	12	3.0
December	3	5	0	2	10	2.5
January	2	4	2	1	9	2.3
February	4	0	2	2	8	2.0
March	1	1	1	2	5	1.3
April	2	1	3	1	7	1.7
May	0	1	1	0	2	0.5
June	0	0	1	1	2	0.5
July	0	0	2	2	4	1.0
August	4	2	3	2	11	2.7
September	3	2	4	3	12	3.0
October	1	2	3	4	10	2.5
November-January	10	9	4	8	31	8.8
February-April	7	2	6	5	20	5.0
May-July	0	1	4	3	8	2.0
August-October	8	6	10	9	33	8.2
Year	25	18	24	25	92	23.0

It will be seen that there is a definite seasonal variation in frequency. The months May-July are least likely to have suitable occasions, and the six months from August to January have nearly three quarters of the total number of occasions.

In the following account, area rainfall will refer to an area mean for a single occasion, and the word "mean" will refer to a mean for a station or area derived from the 92 occasions.

The fall of rain for each station and area were written down for each occasion. From the mean station rainfalls isohyets of average rainfall, representative of low-lying areas, were drawn, and are shown in Fig. 13.

The area rainfalls x_a were then all expressed as ratios of the corresponding mean area rainfalls $\overline{x_a}$, $y_a = (x_a/\overline{x_a})$.

The standard deviation of these ratios is shown in Table 5.

/TABLE 5.

TABLE 5. VALUES OF S_{ya}

Area	S_{ya}
A	0.70
B	0.65
C	0.60
D	0.63
E	0.68
F	0.67
G	0.60
H	0.70
Mean value	0.66

There is no evidence to show that there is any real variation from area to area in the value of S_{ya} and a mean value of 0.66 can be taken.

The correlation coefficients between area rainfalls (ratios of mean) for pair of areas were computed, as also were the serial correlation coefficients for each area of rainfalls on successive occasions. These are shown in Table 6. The serial correlation coefficients run diagonally from top-left to bottom-right.

TABLE 6. CORRELATION AND SERIAL CORRELATION COEFFICIENTS

	A	B	C	D	E	F	G	H
A	-0.12	0.62	0.78	0.42	0.29	0.42	0.26	0.16
B	0.62	0.00	0.74	0.38	0.38	0.52	0.35	0.26
C	0.78	0.74	0.05	0.54	0.39	0.44	0.39	0.17
D	0.42	0.38	0.54	0.11	0.55	0.51	0.51	0.43
E	0.29	0.38	0.39	0.55	0.07	0.68	0.75	0.64
F	0.42	0.52	0.44	0.51	0.68	0.15	0.58	0.59
G	0.26	0.35	0.39	0.51	0.75	0.58	0.04	0.77
H	0.16	0.26	0.17	0.43	0.64	0.59	0.77	0.19

It will be seen that there is little or no serial correlation in any area between successive rainfalls from seeding occasions, and effectively the rainfalls, for any area, may be taken as forming an independent set.

The correlation coefficients between pairs of areas were plotted, a separate map being used for all pairs associated with each area in turn. Isopleths were drawn. The maps are shown as Figs. 5 to 12. They show the way in which the correlation between area rainfalls varies; and they can be used to construct similar maps showing the variation of the correlation coefficient from the centre of any area. To illustrate the method, correlation maps were constructed for the areas V, W, and X (shown in Fig. 4 with pecked lines). For example, for X the correlations of the rainfalls of X with those of the eight standard areas were obtained by interpolation from Figs. 5 - 12. These were plotted on a map for X, and isopleths drawn. Similarly correlation maps were drawn for V and W, and the three estimated maps are given as Figs. 14 to 16.

Estimated correlations of rainfalls for V, X and W with other areas are given in Table 7.

TABLE 7. ESTIMATED CORRELATION COEFFICIENTS

	A	B	C	D	E	F	G	H	V	X	W
V	0.32	0.21	0.29	0.62	0.60	0.52	0.76	0.57	-	0.37	0.63
X	0.57	0.72	0.76	0.53	0.53	0.64	0.42	0.35	0.37	-	0.62
W	0.38	0.46	0.57	0.81	0.70	0.56	0.61	0.47	0.63	0.62	-

Frequency distribution of area rainfalls

The frequency distribution of the ratio y(E) for the area E is given in Table 8.

TABLE 8. FREQUENCY DISTRIBUTION OF y(E)

y(E)	Frequency	y(E)	Frequency
0 - 0.2	4	1.8 - 2.0	2
0.2 - 0.4	10	2.0 - 2.2	1
0.4 - 0.6	16	2.2 - 2.4	2
0.6 - 0.8	17	2.4 - 2.6	2
0.8 - 1.0	12	2.6 - 2.8	2
1.0 - 1.2	8	2.8 - 3.0	1
1.2 - 1.4	5	3.0 - 3.2	0
1.4 - 1.6	3	3.2 - 3.4	1
1.6 - 1.8	6		
		Total	92

The frequency distribution is markedly positive skew. The mode is at about 0.65, and 64 per cent. of the observations have less than the mean value.

The mean values of $x/\bar{x}$ for the separate years 1951, 52, 53, 54 were 0.94, 0.70, 1.18, 1.09.

Regression estimate of y(E)

Let $y = x/\bar{x}$ for the area E, and y_1, y_2, y_3, y_4 be the corresponding ratios for the area G, D, H, F.

Then if $Y = y - \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$

$$S_Y = \alpha \sqrt{\frac{5}{4}(1 - \overline{r_{ij}}) - 2(\overline{r_i} - \overline{r_{ij}})}$$

Then from Table 6

$\overline{r_{ij}} = 1/6\ (0.51 + 0.77 + 0.58 + 0.43 + 0.51 + 0.59)$
$= 0.565$

$\overline{r_i} = 1/4\ (0.75 + 0.55 + 0.64 + 0.68) = 0.655$

and S_Y should thus be given by

$S_Y = (0.66) \sqrt{5/4(1 - 0.565) - 2(0.655 - 0.565)}$

$= 0.398.$

It is of interest to check this from actual values of Y, and also to examine the frequency distribution of Y for nearness to normality. The frequency distribution of Y is given in Table 9.

TABLE 9. FREQUENCY DISTRIBUTION OF Y

Y	Frequency	Y	Frequency	Y	Frequency
-1.1	1	-0.2	13	0.7	1
-1.0	1	-0.1	14	0.8	0
-0.9	1	0	10	0.9	1
-0.8	0	0.1	12	1.0	1
-0.7	0	0.2	5	1.1	1
-0.6	0	0.3	6	1.2	0
-0.5	6	0.4	3	1.3	0
-0.4	4	0.5	3	1.4	1
-0.3	7	0.6	1		
				Total	92

/The

The distribution is symmetrical.

S_Y is found to be 0.395, which agrees with the value 0.398 predicted from the correlation formula.

Table 10 shows the comparison of frequencies with those expected from the corresponding fitted normal curve, fitting mean, standard deviation and total.

TABLE 10. COMPARISON OF FREQUENCY DISTRIBUTION OF Y WITH NORMAL DISTRIBUTION

Y		Frequency	Expected Normal Frequency
	-0.45	9	11.9
-0.45	-0.25	11	12.4
-0.25	-0.15	13	8.5
-0.15	-0.05	14	8.8
-0.05	0.05	10	8.8
0.05	0.15	12	8.8
0.15	0.35	11	15.5
0.35	0.55	6	9.8
0.55		6	7.5
Totals		92	92.0

The value of χ^2, for 6 degrees of freedom, is 10.6, giving P = 0.10. The frequency distribution of Y is thus seen to be reasonably near the normal. We can thus with confidence use the t-test with t given in expression (11).

It was noted that the values of $\frac{1}{4}(y_1 + y_2 + y_3 + y_4)$ differed very little from those of

$$\left[x(G) + x(D) + x(H) + x(F)\right] / \left[\bar{x}(G) + \bar{x}(D) + \bar{x}(H) + \bar{x}(F)\right]$$

The mean of the absolute differences was only 0.03, and so the combined area rainfall of the "control" area (expressed as a ratio of the mean) can be used instead of the arithmetic mean of the area rainfall ratios.

Using the regression equation has given us a residual variable normally distributed with a mean of zero, and with standard deviation 0.39, whereas the "percentage of normal" comparison of y with its "normal" value, 1, gives us a residual variable which is markedly skew, with mean zero and with standard deviation 0.66. The efficiency of the regression technique is further shown in Table 11, which gives the mean values for each year of $y, (y - 1)$, and Y.

TABLE 11. MEAN VALUES FOR AREA E OF y, y - 1, Y FOR EACH OF THE YEARS 1951, 52, 53, 54

Year	y	Mean values of y - 1	Y
1951	0.94	-0.06	-0.02
52	0.70	-0.30	-0.04
53	1.18	0.18	0.08
54	1.09	0.09	-0.04

6. Randomisation

It has been suggested (B. J. Mason, 1955) that with a line of silver iodide generators set up roughly perpendicular to the prevailing wind, a control area, whose rainfall has a high correlation with the seeded area, should be chosen upwind of the generators. Then in order to avoid the effects of weather

/trends,

trends, for example, an abnormally long wet period, seeding should be carried out by random selection on only half the suitable occasions. Rainfall in the seeded area could be predicted from the regression with the control area, and the difference between predicted and actual rainfall could be attributed to seeding on the seeded occasions and to random variations on the non-seeded occasions. Comparison of the two sets of differences would enable the true effects of seeding to be evaluated.

We note that the idea of having control areas upwind of the generators would give very poor correlations. Fig. 17 shows the correlation map for a typical seeded area Q. It will be seen that correlations of rainfall (for our choice of situations) would be 0.5 or less, and these would be valueless. Control areas must be chosen on each side of the seeded area.

Assuming however that this has been done, let us now examine the "randomisation" suggestion. Since there is no serial correlation between successive rainfalls, the "randomisation" method simply doubles the variance of the difference of means that we are testing, and so for two test periods of equal length, one employing "randomisation" and the other not, the "randomisation" method would only be capable of detecting seeding effects $\sqrt{2}$ times as great as those detectable without using randomisation. Putting it another way, the randomisation method would take twice as long to detect a given "seeding" increase as would be needed if all possible seeding occasions were used.

When we used a regression to estimate y for the area E from y_1, y_2, y_3, y_4 for the areas G, D, H, F we found the standard deviation of $Y = y - \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$ to be 0.395. If now we consider the set of Y's to be obtained when seeding took place on the 1st., 3rd., 5th., etc. occasions and no seeding on the 2nd., 4th., 6th., etc. occasions, and we now compute the standard deviation of the quantities $Y_1 - Y_2$, $Y_3 - Y_4$, $Y_5 - Y_6$ etc., we find that it has the value 0.587. The ratio 0.587/0.395 = 1.49, slightly greater than $\sqrt{2}$.

There is no doubt that the randomisation technique is very valuable when there is a high serial correlation among the data being examined; but in the present experiment it is inefficient.

It should in addition be noted that the fear that abnormal weather, for example a long wet period, would lead to spurious results is without justification. For high rainfall in the seeded area would be accompanied by high rainfall in the comparison area with which it has a high correlation, and the difference of the rainfalls (ratios of normal) would, as previously explained, be randomly distributed about a mean of zero. We have already noticed in Table 11 that in 1952, an abnormally dry year (for the seeding occasions), although the seeded area received only 70 per cent. of normal rainfall per occasion, the mean value of the difference was -0.04, merely a small sampling deviation from a true population mean of zero.

7. Detection of seeding increases when the seeded area is unknown

If the seeded area is unknown, the detectable increase due to seeding will be much higher than if we know the area seeded. Practical examples will make this clear.

First example

Let us consider the five areas E, G, D, H, F for the 92 seeding occasions during 1951-54. Assume that one, and only one, of the five areas receives a proportionate increase of P (P = 0, 25, 50, 75, 100) per cent. in rainfall (ratio). The addition was given randomly in turn to the different areas. Then Table 12 shows the number of times that at least

/one

one area now has more than twice normal rainfall when one and only one area, randomly selected, has a percentage increase of rainfall ratio of P. (Curve A of Fig. 18).

TABLE 12. NUMBER OF TIMES FOR GIVEN P THAT AN AREA RAINFALL IS > 200 PER CENT. OF NORMAL. (CURVE A OF FIG. 18)

P	0	25	50	75	100
Number of times in 92	21	26	29	35	42

If we take the value 21/92 to be the mean value of a Poisson variable, then in another set of 92 occasions the probability of having 0, 1, 2, 3 ----- occasions when at least one area has more than 200 per cent. of normal rainfall

$$= e^{-21}\left(1, \frac{21}{1!}, \frac{21^2}{2!}, \frac{21^3}{3!} \ldots\ldots\right)$$

Then we find that the probability p is 0.01 that on 33 or more occasions there would be one or more areas with more than 200 per cent. of normal, and p is 0.001 that on 37 or more occasions there would be one or more areas with more than 200 per cent. of normal.

The frequencies from Table 12 are plotted in Fig. 18, and curve A drawn from them. The lines for p = 0.01 and p = 0.001 are pecked.

Then from curve A we read off that there is a probability of 0.01 that a 64 per cent. increase, given to one and only one area on each occasion, might be undistinguishable from random variations, and at the 0.001 probability level increases of over 82 per cent. would be needed for them to be distinguished from random fluctuations.

Second example

Again consider the areas E, G, D, H, F and give one and one only at random on each occasion a proportionate percentage increase in rainfall ratio of P (P = 0, 25, 50). Write down the differences between the highest and second highest rainfalls, and also the range, on each of the 92 occasions. Then the difference that was exceeded on 21 occasions with 0 per cent. increase was 0.41 and the range that was exceeded on 21 occasions was 1.11.

Table 13 shows the frequencies of difference greater than 0.41 and of range greater than 1.11 for percentage increase P (P = 0, 25, 50).

TABLE 13. (CURVE B OF FIG. 18)

P	0	25	50
Frequency of difference > 0.41	21	28	39
Frequency of range > 1.11	21		38

Curve B of Fig. 18 shows the (common) results of these two methods. From the intersections of the curve with the pecked lines of p = 0.01 and p = 0.001 we see that at the 0.01 and 0.001 probability levels respectively more than 36 per cent. and more than 46 per cent. increases would need to be received by one area to be distinguished from random fluctuations.

These increases are much smaller than those of the first example, which gave 64 and 82 per cent.

/Third

Third example

Consider the five stations in order Manchester, Watnall, Cottesmore, Mildenhall and Felixstowe. Give on each occasion to one and one only, selected at random, of the three middle stations a proportionate percentage increase in rainfall ratio of P (P = 0, 50). Then the difference between the highest of the three middle stations and the lower of the two outer stations is greater than 1.16 on 21 occasions in 92 for P = 0, and on 34 occasions for P = 50. These values are plotted as curve C of Fig. 18. We see that the percentage increases detectable at the 0.01 and 0.001 probability levels are 47 and 60 per cent.

As a check on the reliability of the analysis used in these three examples, let us apply the method to seeding of a known area. Let us consider again the areas E, G, D, H, F. Give on each occasion to the area E a proportionate percentage increase of rainfall P (P = 0, 25, 50). Then compute the values of

$$Y = y - \tfrac{1}{4}(y_1 + y_2 + y_3 + y_4)$$

The value of Y exceeded 0.19 on 21 occasions with 0 per cent. increase in y the rainfall of the area E and exceeded 0.19 on 46 and 76 occasions with percentage increases P = 25 and 50.

These values are plotted as curve D of Fig. 18. We see that the percentage increases required by the area E, to be detected at the 0.01 and 0.001 probability levels, are 13 and 17 per cent. These values are in good agreement with those found by the t-test for the regression technique.

Discussion

In each of the three examples it might be argued that we have been unfair to the comparisons of "unknown location" against "known location", in that we could have a closer network of areas or stations for the "unknown location" methods.

It is, however, unlikely that we could get much improvement for the area network. For a group of areas covering 75 by 75 miles the mean correlation between pairs of areas is about 0.71, and for areas covering 50 by 50 miles about 0.76, and we cannot expect to be able to reduce the detectable increase at the 0.01 level to less than 25 to 30 per cent.

The line method appears to be more capable of improvement, due to closer spacing of stations, when the maximum difference between neighbouring stations is analysed as in the third example. Examination of the correlation charts shows that we can make the assumption that correlation coefficient decrease from 0.9 at 10 miles to 0.9^d at 10d miles distance. It can then be shown that, if there is a close even spacing of stations in a line, with correlation ρ between neighbouring stations, the differences of rainfall amounts between neighbouring pairs of stations form a set which is very nearly independent (very small correlation). Thus the detectable increase will be proportional to $\sqrt{2(1-\rho)}$. In the third example, ρ is about 0.67. The correlation coefficient for two stations, London Airport and Kew, distant six miles from each other, is 0.91. It is doubtful whether we could ever hope to have ρ greater than 0.95, especially since we would not expect seeding effects, if they exist, to have a very sharply defined edge across the plume. Even assuming ρ = 0.95, the method of example three would give a detectable increase of 20 per cent. at the 0.01 probability level. We are, of course, in this latter analysis making the assumption that we can look for seeding effects along a line at a given radial distance from the generators. In

/other

other words, we are not strictly assuming an unknown seeding location.

We must revert to a partial area - type analysis by using two lines of stations, arcs of circles at a radial distance apart of say 30 to 40 miles. With the resulting decrease in efficiency, it is unlikely that we can detect, at the 0.01 probability level, proportionate increases of less than about 25 per cent.

Nevertheless it may be necessary, if we cannot locate the silver iodide plume sufficiently accurately to carry out the comparison between a known seeded area and a chosen control area, to use some sort of "unknown location" comparison. In this event a detailed analysis, probably using the two line method, must be made of historical records of rainfall on "seeded" occasions for stations as closely spaced as possible along the lines, in order to provide an accurate figure for the detectable proportionate increase at the 0.01 probability level, instead of the estimated value of 25 per cent. given above.

8. Distribution of rainfall stations

The area in which we are likely to be interested is England, south of the Humber and east of $2\frac{1}{2}°$W. In this area there are many hundreds of rainfall stations which provide daily readings. But there are seeding occasions when 24-hour falls of rain do not give representative values of rainfall, and 12-hourly readings are more desirable. There are however only some 120 stations from which 12-hourly falls can at present be obtained, and East Anglia and Kent are very poorly represented. It may be possible, with the co-operation of voluntary observers, to have evening readings made at 9 p.m. on days in which precipitation had fallen between 9 a.m. and 9 p.m. This would mean the observer having, on the average, to read in the evening about twice a week. Perhaps some 50 to 100 additional 12-hourly readings might be obtained. For this set of additional stations there would of course be no historical data of 12-hourly falls, but mean 12-hourly falls could be computed by comparison with 24-hour falls, and variability as we have seen does not change much from one part of the area to another; we have also seen that correlations can be computed from area maps.

So no insuperable difficulty is anticipated in making full statistical use of any additional 12-hourly readings we might obtain. If it should be considered undesirable to ask voluntary observers to make the additional readings then we must rely mainly on 24-hourly readings. This will involve some loss of power in the statistical test for detection of seeding increases in rainfall, and our estimated detectable increase, of 13 per cent. in 69 seedings, at the 0.01 probability level must be changed to the order of 16-18 per cent.

In the event of our having to adopt the "unknown location" method, it will obviously be desirable to have a very close network of rain gauges. A good network would be formed by stations along arcs of circles at radii of say 25, 50, 75, 100, 125, 150 miles. With a spacing of stations at 5 miles apart, this would require some 200 stations. If we use 24-hourly readings, this network will be already available. It may however be desirable, even if the location of the plume is known, to set up this network with recording rain gauges if there are encouraging results from the first year of seeding.

9. Conclusions

(i) Rainfall variation in the type of weather situation in which the cloud seeding will be carried out, viz. cloudy (frontal) situations with south-westerly winds, is reduced to a minimum if we restrict the statistical analysis to those occasions with substantial rainfall, defined objectively by an area mean rainfall for the Salisbury area of 5mm.

(ii) To ensure a high seeding efficiency over a substantially wide area, say 20 miles, it is necessary to have a line of some five silver iodide

/generators

generators over a width of 20 to 25 miles. Having a line of generators instead of one only will also greatly increase the probability that we would be able to locate a substantial area which has been efficiently seeded.

(iii) If we can locate such an area on each seeded occasion over a period of three years, then the detectable increase of rainfall due to seeding, measured by the difference between the rainfall in the seeded area and the rainfall in neighbouring non-seeded areas, will be 13 per cent. (at the 0.01 probability level).

If on each occasion we consider the location of the seeded area to be unknown, the detectable increase will be about 25 to 30 per cent.

The respective estimates of 13 per cent. and 25 to 30 per cent. must be increased slightly if we do not obtain additional 12-hourly rainfall readings in East Anglia and Kent.

References

Braham, R.R., B.K. Seely and W.D. Crozier. Trans. Amer. Geophys. Union, 33, 825, 1952.

Johnson, N.L. and H. Tetley. Statistics. Cambridge University Press, 1949.

Jones, R.F. An experiment to test the efficacy of silver iodide seeding in enhancing precipitation. Air Ministry, London, Meteorological Research Papers, No. 876, 1954.

Mason, B.J. Design and evaluation of large-scale rain-making experiments. Nature, Vol. 175, No. 4454, Mar. 12, 1955.

Sutton, O.G. Q.J.R. Met. Soc., Vol. 73, 1947.

Sutton, O.G. Micrometeorology. McGraw Hill, London. 1953, pp.287, 292.

Vittori, O. Sur la détermination de la distribution dans l'atmosphère des noyaux de condensation et de particles obtenues par sublimation. Rome, 1954. (Reviewed in Bull de l'Obs. du Puy de Dome, 1955, No. 1).

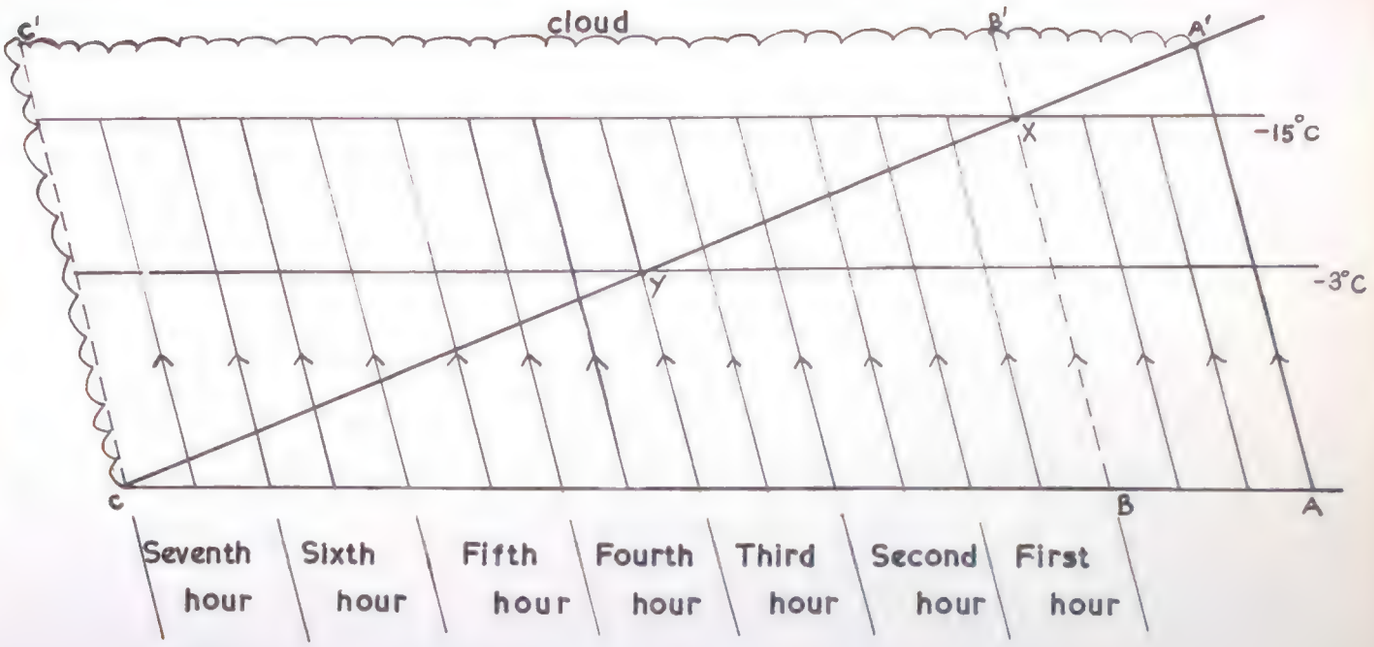


FIG. 1 Vertical diffusion of AgI.

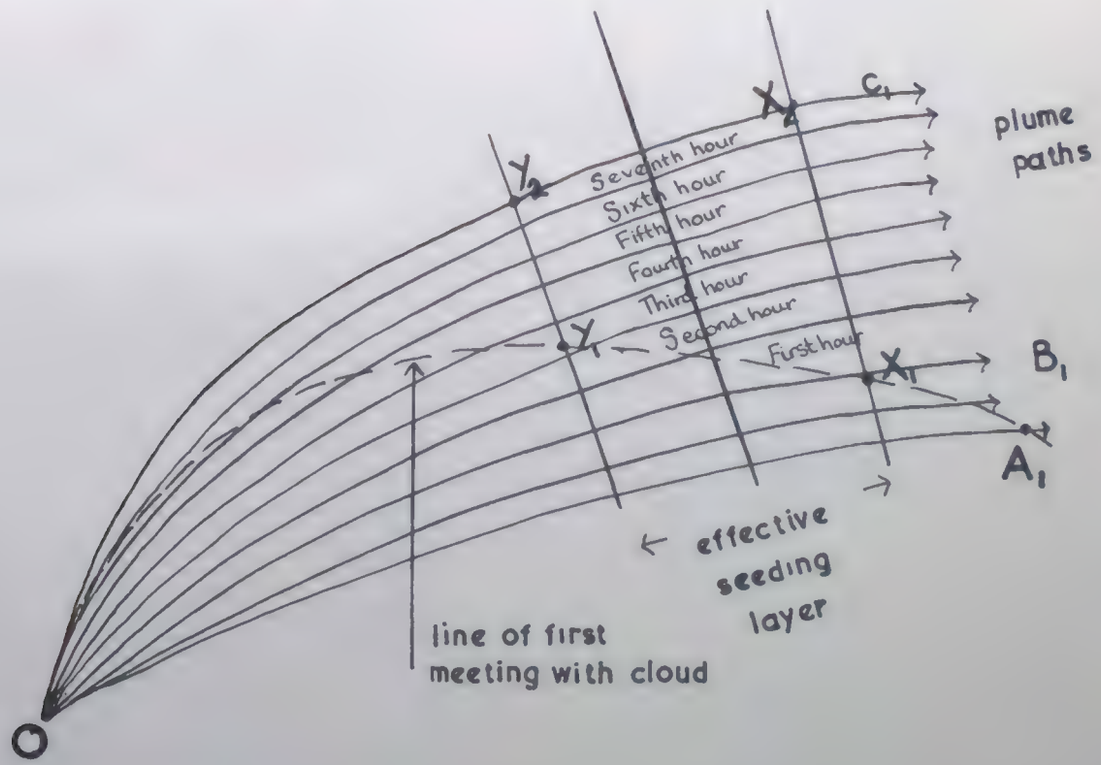
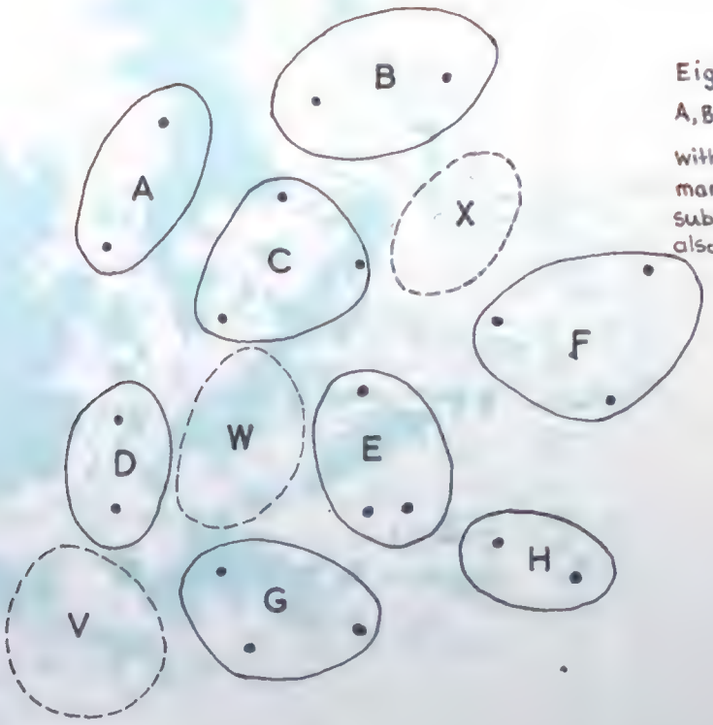


FIG. 2 Horizontal projection of particle paths of AgI

		7
	7	6+7
7	6+7	5+6+7
6+7	5+6+7	4+5+6+7
5+6	4+5+6+7	3+4+5+6+7
4+5	3+4+5+6	2+3+4+5+6+7
3+4	2+3+4+5	1+2+3+4+5+6
2+3	1+2+3+4	1+2+3+4+5
1+2	1+2+3	1+2+3+4
1	1+2	1+2+3
	1	1+2
		1
10 miles	20 miles	30 miles

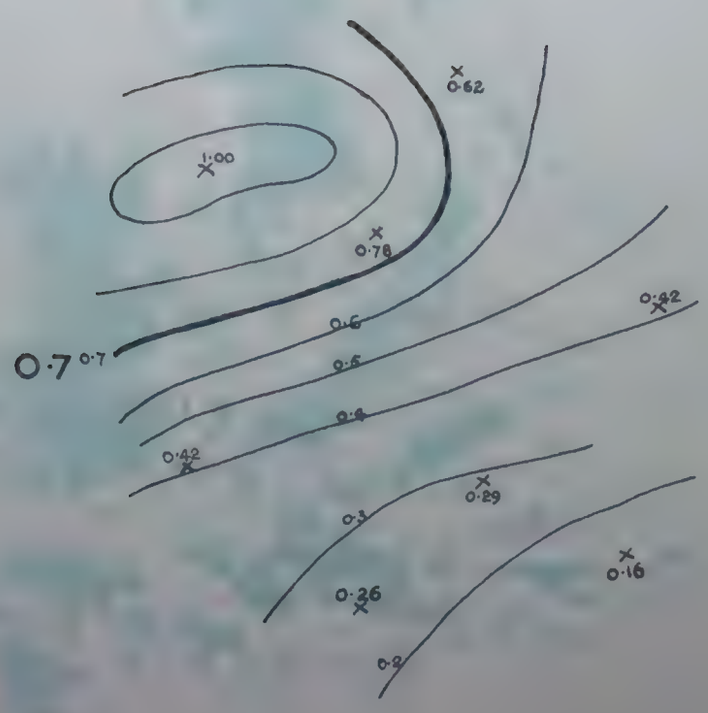
FIG. 3 30 miles latitudinal movement of plume, with plume widths of 10, 20 and 30 miles

FIG. 4



Eight sub-areas
A, B, C, D, E, F, G, H
with rainfall stations marked. Additional sub-areas V, W, X are also marked.

FIG. 5



A correlations
With other areas

M.R.P. 931

FIG. 6

B correlations with other areas

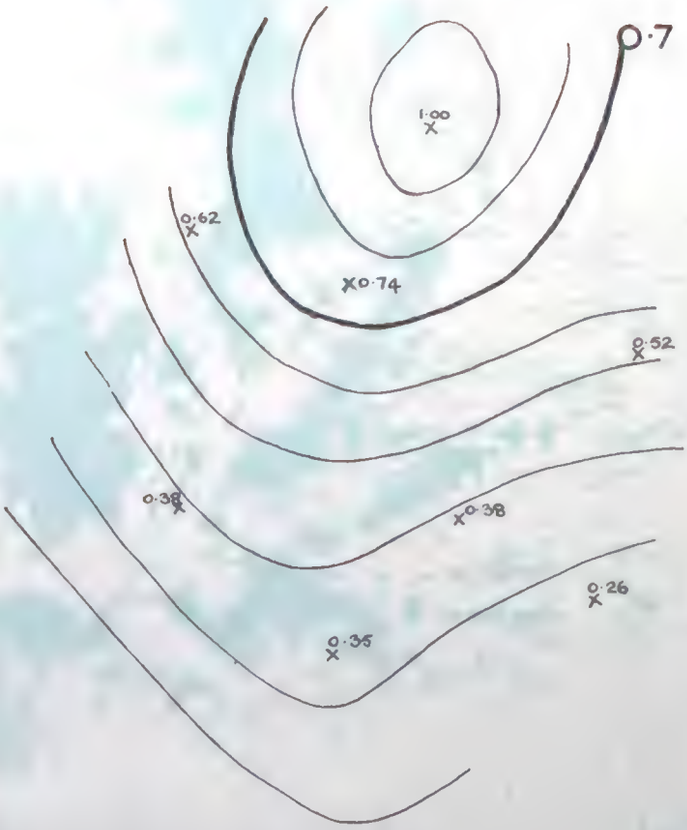


FIG. 7

C correlations with other areas

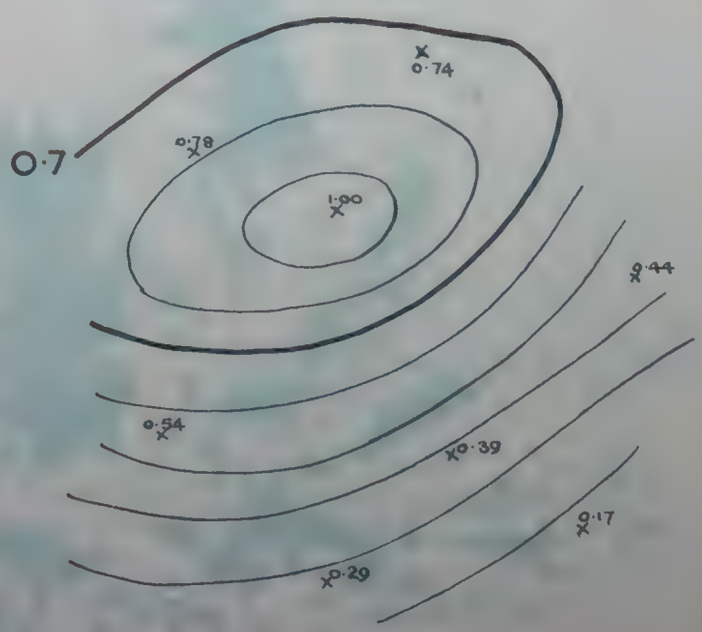


FIG. 8

D correlations with other areas

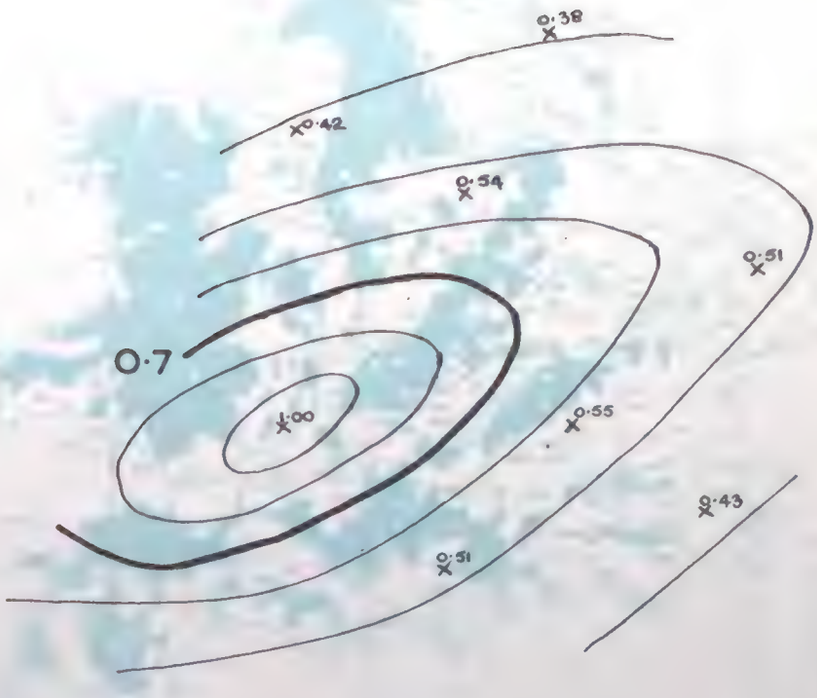


FIG. 9

E correlations with other areas

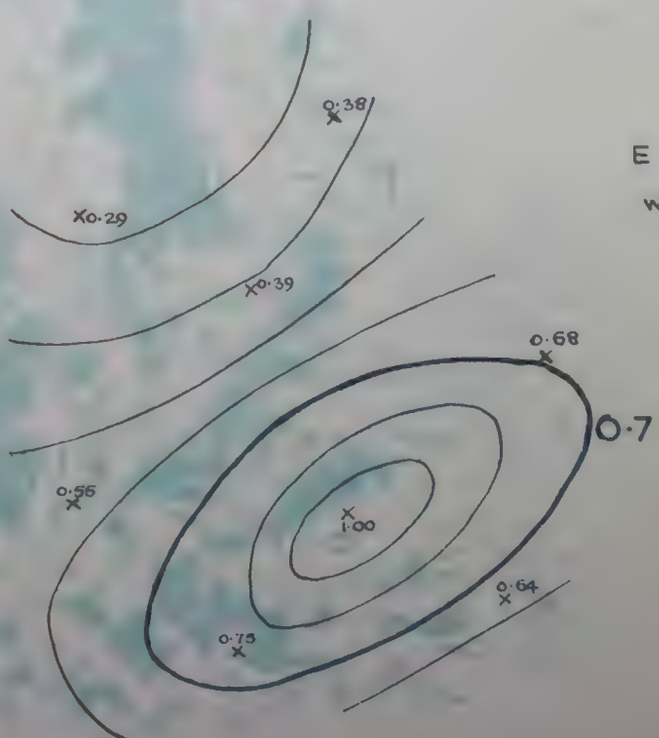


FIG. 10

F correlations with other areas

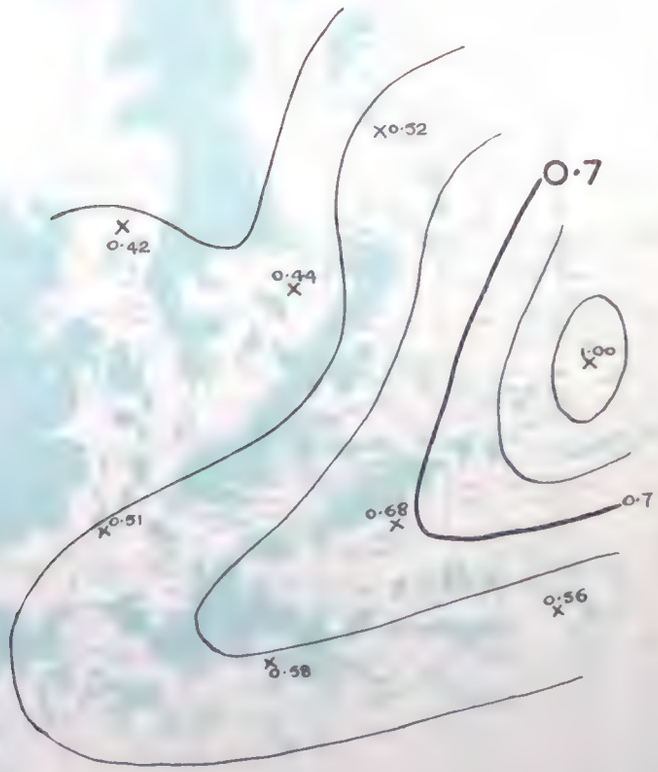


FIG. 11

G correlations with other areas

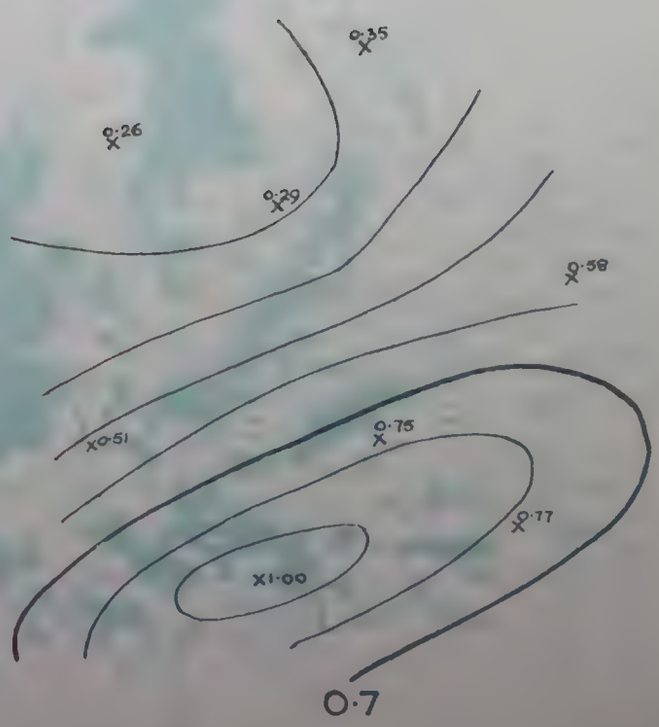


FIG. 12

H correlations with other areas

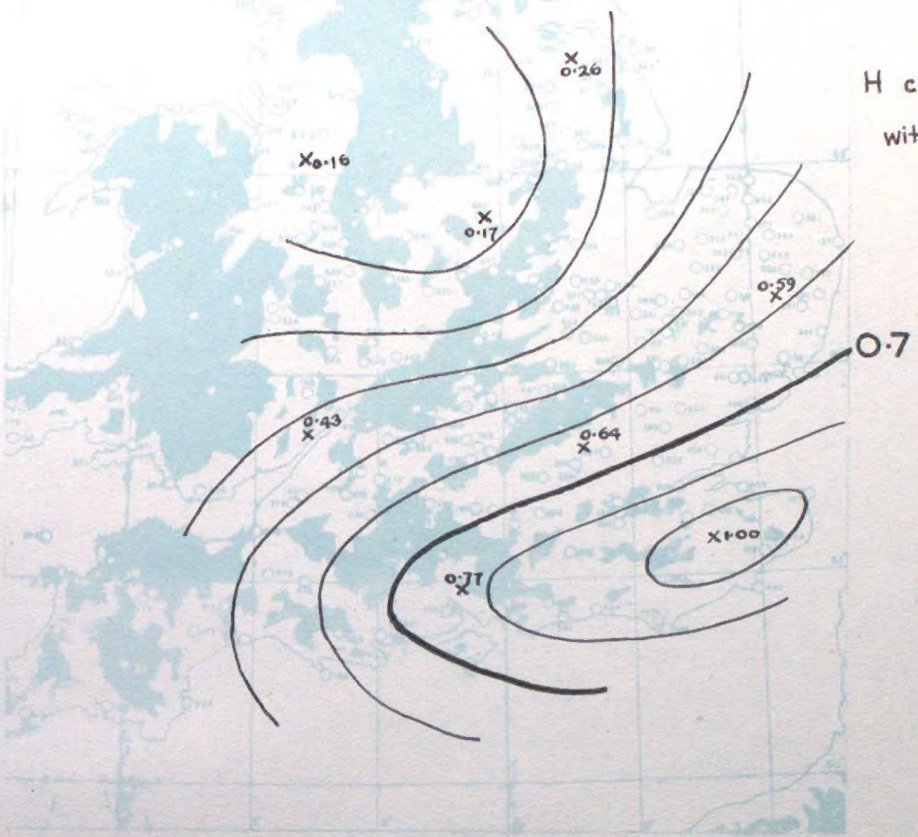


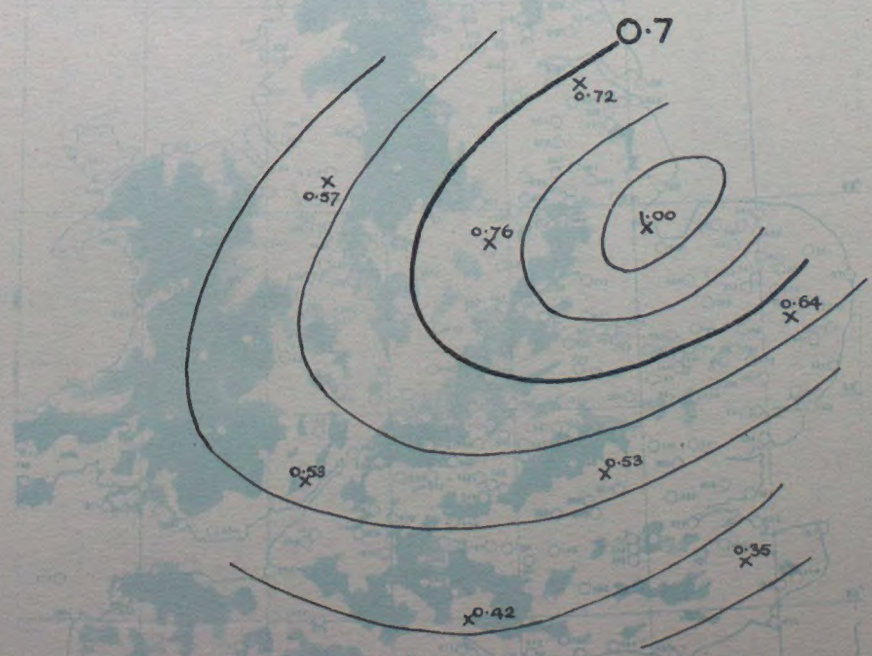
FIG. 13

Average rainfall amount (mm.) from 92 occasions 1951 - 1954

FIG. 14

V estimated correlations with other areas

FIG. 15



X estimated correlations with other areas

FIG. 16

W estimated correlations with other areas

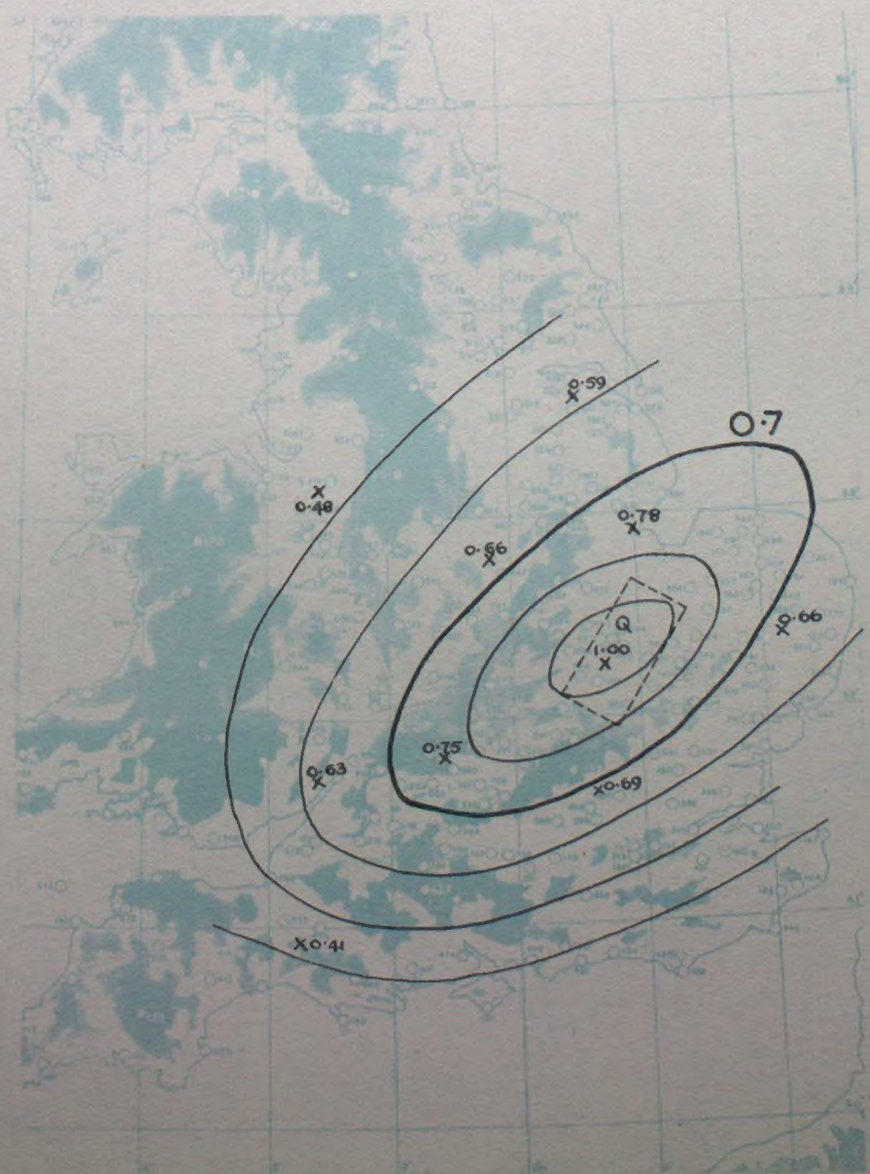
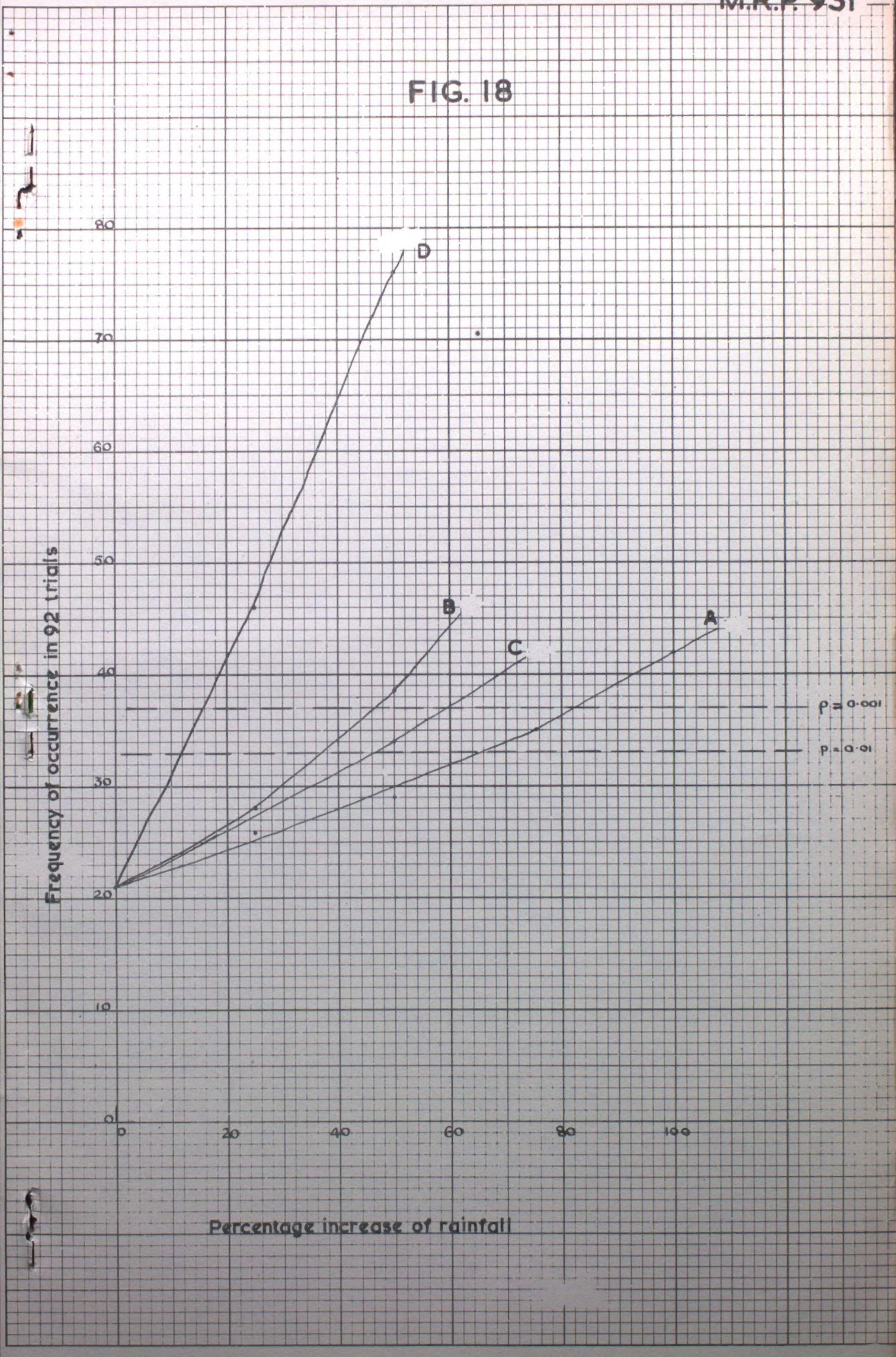


FIG. 17

Typical "seeded" area with correlations

FIG. 18


80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
D
B
C
A
p = 0·001
p = 0·01
Frequency of occurrence in 92 trials
Percentage increase of rainfall